

On yield, fault distributions, and clustering of particles

by C. H. Stapper

Increasing the levels of semiconductor integration to larger chips with more transistors causes the fault and defect distributions of VLSI memory chips to deviate increasingly further from simple random Poisson statistics. The spatial distributions of particles on semiconductor wafers have been analyzed to gain insight into the nature of integrated circuit defect statistics. The analysis was done using grids of squares known as quadrats. It was found that the cluster parameter, which until now has been treated as a constant, did vary with quadrat area. The results also show that the deviation from Poisson statistics continues to increase into the realm of wafer-scale integration or WSI. Computer simulations were used to verify this effect.

Introduction

From the onset of integrated circuit development in 1964 it was realized that simple Poisson statistics was not appropriate for the modeling of integrated circuit yield calculations [1]. Originally this was looked upon as an aberration caused by high defect levels and poor control during the early stages of manufacturing. Conventional wisdom of the early 1970s took for granted that the lowering of defect levels and improved manufacturing cleanliness

would eventually result in pure randomness of defects. It was believed that low defect densities could therefore be modeled with the random defect statistics associated with Poisson's distribution. Similarly, some theoreticians claim that for large regions on wafers and full wafers the defect statistics should also revert to pure simple Poisson statistics.

It is the experience of this author, however, that integrated circuit manufacturing lines are more perverse. For instance, it has already been reported that a reduction of defect levels results in an increased deviation from Poisson statistics [2]. The same effect has now also been observed as a result of increased chip area, circuit complexity, and both of these combined. An investigation into the nature of this effect is the subject of this paper.

The deviation from Poisson statistics manifests itself by producing very long tails in the distributions of the number of faults per chip. A fault is defined as a defect which causes a chip failure. An example of fault distributions can be seen in Figure 1, which shows data from 450 experimental dynamic random access memory chips. These chips were processed under very clean conditions. The data in Figure 1 were obtained with the array diagnostic technique described by Gangatirkar, Presson, and Rosner [3]. Shown are the frequency and cumulative distributions of the number of failing single cells, failing single rows or word lines, and failing single columns or bit lines. Not included here are some other faults such as the ones that result in adjacent cell, row, or column failures and defects that cause large sections of chips or entire chips to fail.

The graphs on the left-hand side in Figure 1 represent the frequency distributions of the faults. These data have been truncated and the remaining cells have been added to the buckets of 40 faults per chip. This affects primarily the

©Copyright 1986 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

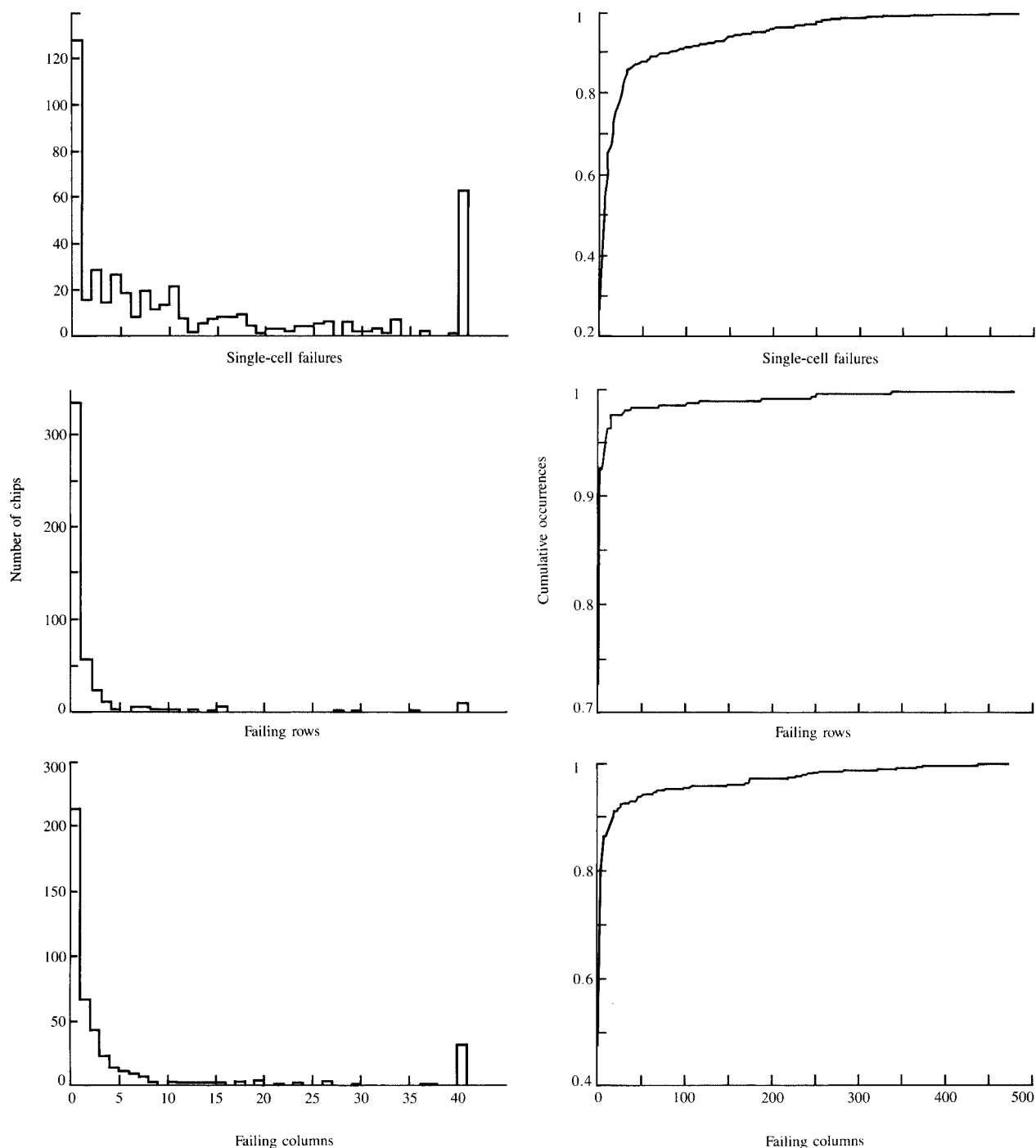


Figure 1

Fault distributions: The graph indicates the relative number of failing single cells, failing rows, and failing columns that were observed in 1M DRAM chips.

distributions of the single cell and column failures. On the right-hand side of Figure 1 the cumulative distributions are shown. These have been extended to include the data points

with more than 40 failures per chip. The tails of these distributions appear well-behaved and include very large numbers of failures for all three cases.

Comparing data to Poisson statistics

Some interesting statistics are associated with the results in Figure 1. The data on the extreme left in the graphs represent the number of chips that did not fail electrical test. These chips are fault-free and therefore represent the yield. The actual values are 27.4% for the single cells, 72.5% for the single rows, and 47.3% for the single columns. If these yields had been associated with pure Poisson random defect statistics, they would be related to the average number of faults per chip by

$$Y = e^{-\lambda} \quad (1)$$

In this equation Y represents the yield and λ the average number of faults per chip. Equation (1) can be inverted to give

$$\lambda = -\ln Y \quad (2)$$

thus making it possible to calculate the average number of faults per chip from the yield. For the yields in Figure 1 such calculations result in an average of 1.296 single cells, 0.321 single rows, and 0.748 single columns per chip. The actual averages for the distributions in Figure 1 are 28.65 single cells, 4.51 single rows, and 15.37 single bit lines. These values are an order of magnitude higher than those calculated with Equation (2). It can therefore be concluded that Poisson statistics do not provide a good model for these data; namely, defect levels calculated with those statistics are too low.

That Poisson statistics are not applicable to the data of Figure 1 can also be demonstrated in another way. Poisson's distribution is given by the formula

$$P(X = k) = e^{-\lambda} \lambda^k / k! \quad (3)$$

where X is a random variable designating the number of faults per chip and k an integer having values of 0, 1, 2, 3, etc. It is generally known, and not difficult to prove, that the mean and variance of Equation (3) are both given by the value of λ . In the distributions of Figure 1 the variances are 4299.9, 932.4, and 3285.1 for the cells, rows, and columns respectively. These values are at least two orders of magnitude higher than the average number of faults. The statistics needed to model these distributions are therefore expected to be more complex than Poisson statistics. In the next sections some possible causes for this complexity are investigated.

The deviation from Poisson's distribution can be quantified with the ratio of the variance and the mean. For a Poisson distribution this ratio is equal to one. For the single-cell, single-row, and single-column distributions of Figure 1 these ratios become 150.1, 206.7, and 213.7. These are the highest values of this ratio that this author has ever observed for fault distributions of integrated circuit chips. These data, however, also came from chips with the largest physical area, the greatest number of transistors, the smallest feature sizes,

and the thinnest dielectrics ever analyzed in this way by this author. These data led to the investigation of the cause for this deviation from Poisson's distribution in very large and complex chips. Some of the early results of this study are the subjects discussed in the following sections of this paper.

Compound Poisson statistics

That simple Poisson statistics could not be used for integrated circuit yield calculations was confirmed experimentally in the 1960s. It was clear from the data that this effect was caused by defect clustering [4-7]. It was also determined in the early 1970s that the wafer-to-wafer variation of defect densities could be responsible for the same effect [8]. In either case, the data could only be modeled with some form of mixed or compound Poisson statistics. The nature of these statistics is discussed in this section.

The compounding or mixing process can be applied to Poisson's distribution (3) by assuming the value of λ to be another random variable. This can be justified by assuming that there are a number of independent regions, each having random faults, with a different average number of faults per unit area. If each region is designated by an index number i , then the corresponding average number of faults is indicated by λ_i . Within these regions Poisson's distribution is assumed to be valid with λ_i as a parameter. Associated with the values of i are a set of probabilities P_i , which have to satisfy the normalization condition

$$\sum_{i=0}^{\infty} P_i = 1 \quad (4)$$

This normalization assumes that there are an infinite number of regions and values of λ_i . Situations with fewer regions can be modeled by assigning a probability $P_i = 0$ to any excess regions.

The mean or average number of faults associated with the discrete probability distribution P_i is given by

$$E(\lambda) = \sum_{i=0}^{\infty} \lambda_i P_i \quad (5)$$

Similarly, the variance is given by

$$V(\lambda) = \sum_{i=0}^{\infty} \lambda_i^2 P_i - E(\lambda)^2 \quad (6)$$

The probabilities P_i can be used to compound Poisson's distribution (3) by the summation

$$P(x = k) = \sum_{i=1}^{\infty} P_i e^{-\lambda_i} (\lambda_i)^k / k! \quad (7)$$

This is the general form of the mixed or compound Poisson distribution that results from discrete compounding.

It is also possible to use a continuous probability distribution function $P(\lambda)$ for compounding. A detailed

derivation of this technique is described in [8]. It should be noted here that the normalization requires

$$\int_0^{\infty} P(\lambda) d\lambda = 1. \quad (8)$$

The mean and variance of $P(\lambda)$ can be calculated with

$$E(\lambda) = \int_0^{\infty} \lambda P(\lambda) d\lambda, \quad (9)$$

and

$$V(\lambda) = \int_0^{\infty} \lambda^2 P(\lambda) d\lambda - E(\lambda)^2. \quad (10)$$

Compounding of Poisson's distribution (3) with the probability distribution function $P(\lambda)$ is done with the integral

$$P(X = k) = \int_0^{\infty} d\lambda P(\lambda) e^{-\lambda} \lambda^k / k!. \quad (11)$$

This is the fault distribution resulting from a continuous compounding or mixing process.

It is not difficult to prove that the mean and variance for the fault distributions formulated in Equations (7) and (11) are given by

$$E(X) = E(\lambda) \quad (12)$$

and

$$V(X) = E(\lambda) + V(\lambda). \quad (13)$$

This last equation shows why compound Poisson statistics is useful for modeling the fault distributions of integrated circuits: The variance in Equation (13) is always greater than the mean in Equation (12). The ratio of the variance and the mean for distributions (7) and (11) is given by

$$\frac{V(X)}{E(X)} = 1 + \frac{V(\lambda)}{E(\lambda)}, \quad (14)$$

which is always greater than one.

The ratio of the variance and the mean can be better understood by assuming the existence of a complexity factor. This factor depends on chip areas, ground rules, or feature sizes of the photolithographic patterns, manufacturing complexities, dielectric thicknesses, circuit design sensitivities, and a host of unknown factors. It may therefore be somewhat presumptuous to assume that in this case the average number of faults for any chip can be expressed by

$$\bar{\lambda} = C\bar{\lambda}_0, \quad (15)$$

where C is the complexity factor and $\bar{\lambda}_0$ the average number of faults on a chip with unit complexity. However, if this assumption is made, it would imply that in general

$$\lambda = C\lambda_0, \quad (16)$$

where λ_0 is a random variable distributed in the same way as λ . It follows therefore that

$$E(\lambda) = CE(\lambda_0) \quad (17)$$

and

$$V(\lambda) = C^2 V(\lambda_0). \quad (18)$$

As a result the ratio of $V(X)/E(X)$ in Equation (14) becomes

$$\frac{V(X)}{E(X)} = 1 + C \frac{V(\lambda_0)}{E(\lambda_0)}. \quad (19)$$

This implies that the ratio of the variance and the mean of integrated circuit fault distributions increases linearly with the complexity factor C . It also explains the observations that were described in the previous sections.

The complexities of modern integrated circuits make it—unfortunately—impossible to define a single complexity factor for each chip. The effect of complexity on the pertinent statistics has therefore been studied with particle distributions on blank integrated circuit wafers. This eliminates the presence of photolithographic patterns and circuit sensitivities. As a consequence, the complexity factor reduces to an area ratio. The results of the studies are therefore reported in terms of areas or relative areas in subsequent sections and figures in this paper.

It is possible to define some ratios related to the distributions of λ that are independent of the complexity factor. One of these is the coefficient of variation, which is defined by the ratio of the standard deviation and the mean. It is denoted here by σ/μ and is calculated with the formula

$$\sigma/\mu = \sqrt{V(\lambda)}/E(\lambda). \quad (20)$$

Introduction of relationships (17) and (18) into this equation gives

$$\sigma/\mu = \sqrt{V(\lambda_0)}/E(\lambda_0), \quad (21)$$

which is indeed independent of the complexity factor C .

In the following sections a cluster parameter α is used. This quantity is related to the coefficient of variation by

$$\alpha = \sqrt{\mu/\sigma}. \quad (22)$$

It is therefore also independent of the complexity factor. For the distributions resulting from Equations (7) and (11) it is possible to define

$$\alpha = \frac{E(\lambda)^2}{V(\lambda)} \quad (23)$$

as a general quantity. Its dependence on area, and therefore chip complexity, is also studied in the following sections.

The preceding discussion was completely general. The type of compounders required in Equations (7) and (11) depends entirely on the nature of the fault distributions observed on manufactured chips. Until now the use of a gamma distribution in Equation (11) has been found useful;

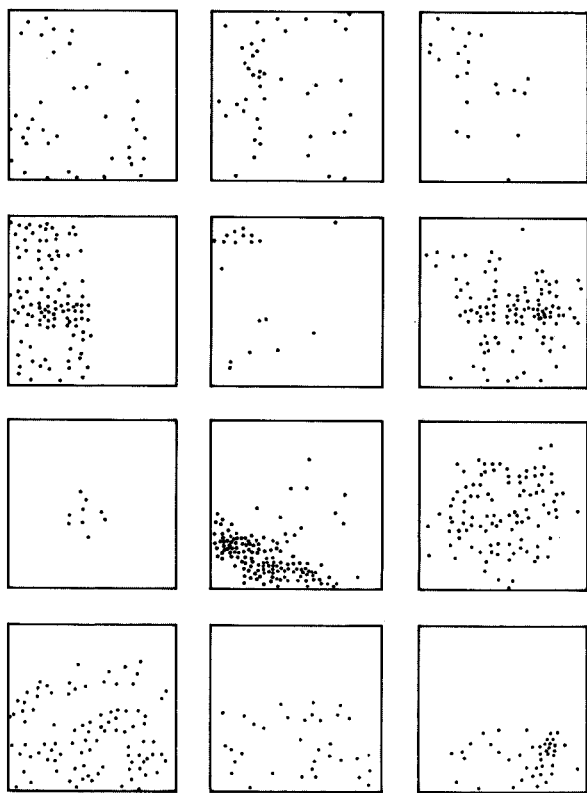


Figure 2

Particle maps: The locations of particles on wafers obtained with an electronic particle detector.

it has been reported in [8–11]. The properties of this distribution and some other compound distributions are given in the Appendix.

The effect of defect clustering

The clustering of defects on wafers in principle should not be difficult to analyze. Wafer maps showing the location of usable and failing chips can be obtained by anyone working in the industry. However, finding the number and the location of the failing defects on nonfunctioning chips is not easy. Until now, only data for half a wafer have appeared in the literature. These were results published by Moore [6] that have been extensively analyzed by Warner [7, 12] and Stapper [9]. Moore's data do produce a long tail in the fault-per-chip distribution. However, the ratio of the variance to the mean of these data was only 3.94.

In order to get more insight into the statistics of clustering, F. M. Armstrong and K. Saji [13] decided to analyze particle distributions on blank wafers from a manufacturing line. They reasoned that some of these particles might result in defects that cause chip failures. Understanding the nature of

these particle distributions could therefore give an insight into the statistics applicable to integrated circuit yields.

The spatial distributions that Armstrong and Saji investigated were obtained with an electronic particle detector. This tool used scattered light to pinpoint the location of particles on wafer surfaces, and was used to analyze equal square areas on twelve wafers. This resulted in the maps of particle locations that are shown in Figure 2. Armstrong and Saji subdivided each one of these maps into 36 smaller squares or "quadrats." (The word *quadrat* originates from the methods used by ecologists to analyze the spatial population distributions of animal and plant life. A. Rogers [14] applied this technique to study the spatial distributions of retail stores in cities. Armstrong and Saji used the "quadrat method" for the analysis of particles on wafers.) They then counted the number of particles in each quadrat and thus determined the frequency distribution of the number of particles per quadrat. These data were collected individually for each wafer.

The results of the particle counting were analyzed with a method described by Rogers [14]. This consisted of determining which one of four different compound distributions provided a best fit to these data. In order to do this, the parameters for the theoretical distributions were calculated using a maximum likelihood technique. The goodness of fit for the result was established with a chi-square test. In this way it was found that the distributions from four wafers were best modeled with a mixed Poisson-binomial distribution, four others with a Neymann Type A distribution, and three with a negative binomial distribution. For one wafer all three of these distributions fitted equally well. In all cases it was found that any one of these compound distributions gave a much better fit to the data than Poisson's distribution. The formulas and properties of the four distributions that were used in this study are described in the Appendix of this paper.

Armstrong and Saji also studied the fraction of quadrats without any particles. This fraction was considered to be the quadrat yield. They found that the smallest difference between the theoretical yield and the observed yield occurred with Neymann Type A statistics on eight wafers, negative-binomial statistics on three wafers, and Poisson-binomial statistics on one wafer.

Results like these depend on the conditions encountered by the wafers in the factory. Although such conditions can be expected to differ between factories, this author anticipates that particle clustering is universal in the industry. As a consequence, some form of compound statistics is expected to be applicable for yield models everywhere.

Only a few of the particle maps in Figure 2 produced particle distributions with long tails when analyzed by Armstrong and Saji. This suggests that defect clustering could indeed be a cause for the tails of the fault distributions

in Figure 1. However, since the majority of the maps in Figure 2 did not show this behavior, there must be an additional contributor. This is the subject of the next section.

Wafer-to-wafer variations

It has been described in [8] how compound Poisson statistics can also result from wafer-to-wafer variations in defect levels. It is not difficult to determine whether this is the case with particles on wafers. **Figure 3** shows the cumulative distribution of the number of particles per map for the twelve maps in Figure 2. The result from this small sample is rather scattered, and it is difficult to draw any definitive conclusions about the nature of the number of particles per wafer distribution. Nevertheless, these data have a mean of 62.75 and a variance of 1775.5, which results in a variance-to-mean ratio of 28.29. Simple Poisson statistics does not therefore appear to be applicable. Some form of compound Poisson statistics may be more applicable for modeling these results.

The preceding experiment is not difficult to repeat. At IBM, cleaned blank wafers are periodically sent through various tools used for photoresist application, photolithographic exposure, pattern etching, evaporation, etc. The number of particles per wafer is counted before the wafer enters the tool and again after it exits. A cumulative distribution of the number of particles observed on 167 of such wafers is shown in **Figure 4**. These data have a mean of 31.75 and a variance of 4116.2. The ratio of variance to the mean is therefore 129.6, which is higher than the ratio of 28.3 in the Armstrong-Saji data. This indicates that the map-to-map variation of the number of particles per map in Figure 3 is less than the wafer-to-wafer variation of particles found in a run-of-the-mill sample. An analysis of the Armstrong-Saji data with respect to the map-to-map variation is therefore expected to be conservative. This is one of the reasons why these maps are studied in detail in this paper.

Armstrong and Saji analyzed the particle distributions individually for each one of the maps in Figure 2. They used a single grid with a single quadrat area. In the analysis made in this paper, the data for all the maps are combined and six different quadrat sizes are used. This makes it possible to study the applicable statistics and distributions as a function of quadrat size. Furthermore, since no photolithographic patterns or circuit sensitivities are present, the previously described complexity factor is simply equal to the quadrat area.

The smallest quadrat area that was studied was obtained by dividing the maps in Figure 2 into 12×12 squares. This produced 144 quadrats per map. The next larger size used an 8×8 grid with 64 squares. The third size consisted of 6×6 squares, for a total of 36 per map, and is the same quadrat grid used by Armstrong and Saji. The other quadrat arrangements that are analyzed here are 4×4 , 3×3 , and 2×2 , each with 16, 9, and 4 squares per map respectively.

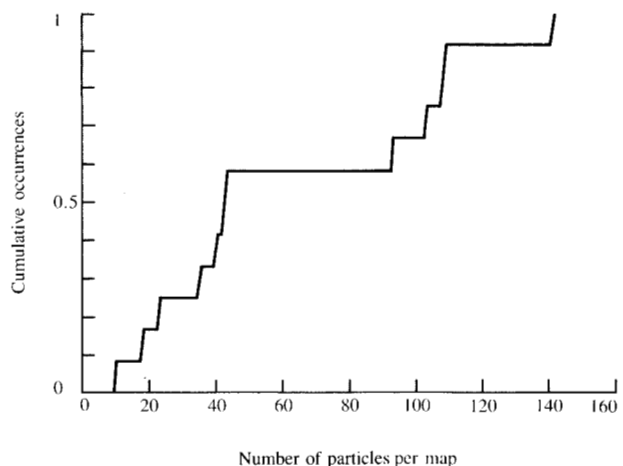


Figure 3

A cumulative particle distribution: This curve results from the number of particles on each map in Figure 2.

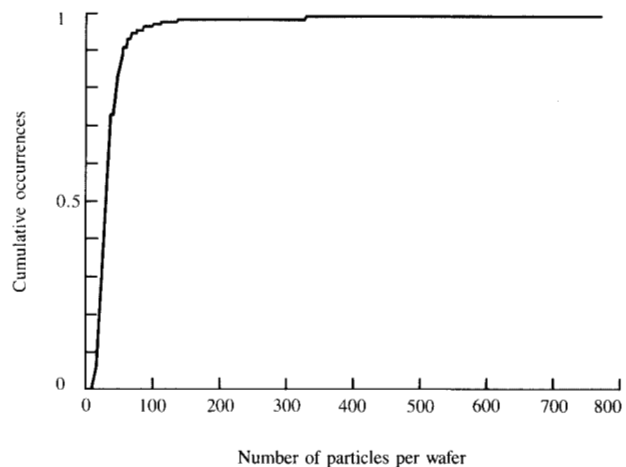


Figure 4

Particle distributions from a manufacturing line: The cumulative distribution of the number of particles per wafer. These numbers were counted on wafers that were sent through manufacturing tools to determine their cleanliness.

The distributions of the numbers of particles per quadrat resulting from this analysis are tabulated in **Table 1**. It is clear from the result that the distributions for the 4×4 , 3×3 , and 2×2 quadrat arrangements do have long tails. The mean, variance, and mean-to-variance ratios are given in **Table 2**. The results show that the larger quadrats have the greatest deviation from Poisson statistics. This confirms the earlier observations and indicates that this effect could have been caused by particle clustering and the wafer-to-wafer variation of clustered particles.

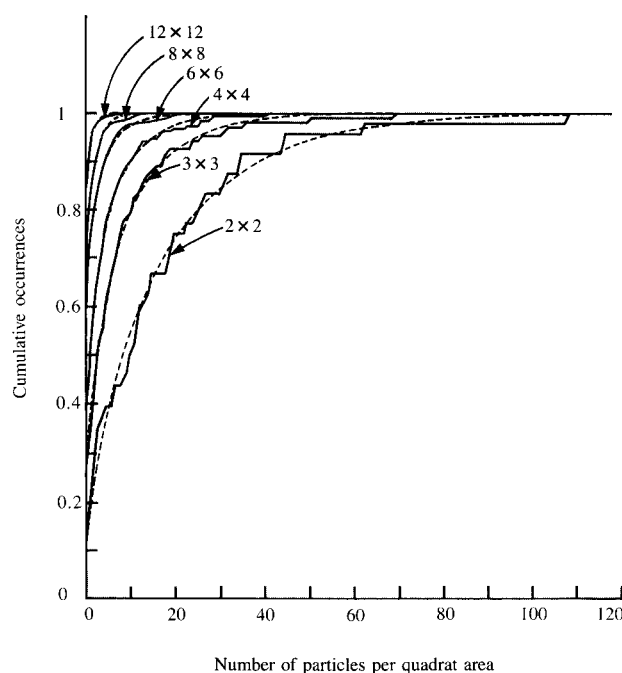


Figure 5

Distributions of the number of particles per quadrat: Data and models for the cumulative distributions of the number of particles per quadrat. These results were obtained from the maps in Figure 2.

Statistical models

It is not difficult to fit the data in Table 1 with compound Poisson statistics such as Neymann Type A, negative binomial, Poisson-binomial, or the Poisson-negative binomial, as described by Rogers [14]. For the analysis used in this paper only negative binomial statistics were used, since they appeared to provide an entirely satisfactory model. This is shown in Figure 5, which represents several cumulative distributions of the number of particles per quadrat. Each graph shows the data and the theoretical results corresponding to a different quadrat area. The solid lines represent the data in Table 1; the dashed lines are cumulative negative binomial distributions calculated with the formula

$$C(x) = \sum_{k=0}^x \frac{\Gamma(\alpha + k)}{k! \Gamma(\alpha)} \frac{(\lambda/\alpha)^k}{(1 + \lambda/\alpha)^{\alpha+k}}. \quad (24)$$

In this expression λ is the mean number of particles per quadrat in each distribution, α is a clustering parameter, and x is the cumulative number of particles per quadrat given by an integer 0, 1, 2, etc. The parameters λ and α were determined with a nonlinear least-square method using Equation (24) and the cumulative distribution corresponding to the data in Table 1. The parameter values obtained in this way are tabulated in Table 3.

Two important conclusions can be drawn from these results. Both conclusions have to do with a model described in [15]. In that paper it was claimed that the negative binomial distribution and the yield are affected by an experimental dependence of λ on area. This dependence is plotted in Figure 6. Also shown is the straight line that corresponds to

$$\lambda = AD, \quad (25)$$

where A is the quadrat area and D the average number of particles per unit area. When the quadrat area of the 144×144 grid arrangement is taken as unity, the defect density D is equal to 0.4358 particles per unit area. It is reassuring to find that this line corresponds closely to the data points. The experimentally determined parameter λ does therefore represent the average.

Table 1 Distributions of the number of particles per quadrat observed for different quadrat areas.

Particles per quadrat	Quadrat arrangement					
	12 × 12	8 × 8	6 × 6	4 × 4	3 × 3	2 × 2
0	1275	459	229	62	28	5
1	293	138	67	32	6	4
2	101	83	39	21	10	5
3	29	32	25	12	13	3
4	8	23	23	10	3	1
5	11	14	11	13	8	1
6	3	1	10	4	4	0
7	1	4	8	7	5	2
8	5	0	6	5	4	0
9	1	3	3	4	3	1
10	1	2	1	3	1	2
11		3	0	1	4	1
12		3	1	4	1	3
13		1	0	3	2	1
14		2	1	0	2	1
15			1	1	1	2
16			0	0	1	0
17			1	2	0	0
18			1	1	3	0
19			1	0	1	2
20			2	1	0	2
21			0	0	0	0
22			1	0	0	0
23			0	1	0	1
24			0	0	2	0
25			0	0	0	1
26			0	2	1	1
27			0	0	0	1
28			1	1	0	0
29				1	0	0
30				0	0	0
31				0	1	1
32				0	1	1
33				0	0	0
34				0	0	0
35				0	0	2
36				0	1	0
37				0	0	0
38				0	0	0
39				0	0	0
≥40				1	2	4

It was also suggested in [15] that the components of the cluster parameter α might be dependent on the area of a chip. As a result α would become a function of chip area. In Figure 7 the experimentally determined values of α are plotted as a function of quadrat area. This shows that α is indeed dependent on these areas. The results also indicate that these values have a minimum for the quadrats of the 6×6 arrangement. Low values of α in the negative binomial distribution are associated with high degrees of clustering. Maximum clustering appears therefore to be taking place for these quadrat areas.

Yield as a function of quadrat area

Yield as a function of integrated circuit chip areas has been at the center of yield modeling ever since the publication of [1]. In that paper it was suggested that a plot of the logarithm of yield versus chip area should curve upwards. We can now see if this holds for the data in Figure 2 by defining the yield to be equal to the fraction of particle-free quadrats. The result of this yield as a function of relative quadrat area is shown in Figure 8. The data points do suggest upwards curvature as predicted in [1].

It is possible to approximate the data in Figure 8 with the negative binomial yield model of [6] in the form

$$Y = Y_0(1 + AD/\alpha)^{-\alpha}, \quad (26)$$

where Y_0 is a gross particle yield, A the quadrat area, D the particle density per unit area, and α a constant cluster

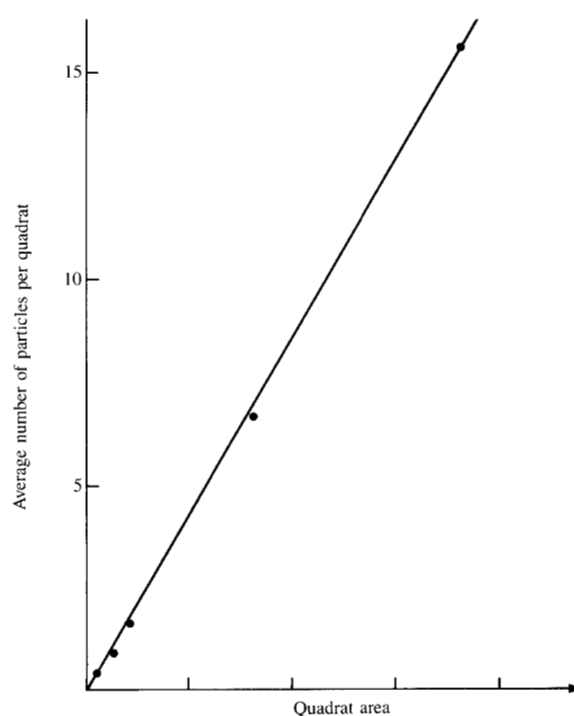


Figure 6

Number of particles vs quadrat area: The average number of particles per quadrat depends linearly on quadrat area.

Table 2 Means and variances of the distributions in Table 1.

	Quadrat arrangement					
	12×12	8×8	6×6	4×4	3×3	2×2
Relative area	1	2.25	4	9	16	36
Mean	0.4358	0.981	1.74	3.92	6.97	15.69
Variance	0.9785	3.605	11.26	37.43	110.5	384.2
Variance-to-mean ratio	2.25	3.68	6.45	9.55	15.84	24.49

Table 3 Model parameters resulting from a least-square analysis.

	Quadrat arrangement					
	12×12	8×8	6×6	4×4	3×3	2×2
Relative area	1	2.25	4	9	16	36
Parameters						
λ	0.4191	0.9418	1.691	3.829	6.671	15.536
α	0.4896	0.4637	0.3757	0.5010	0.5824	0.7301

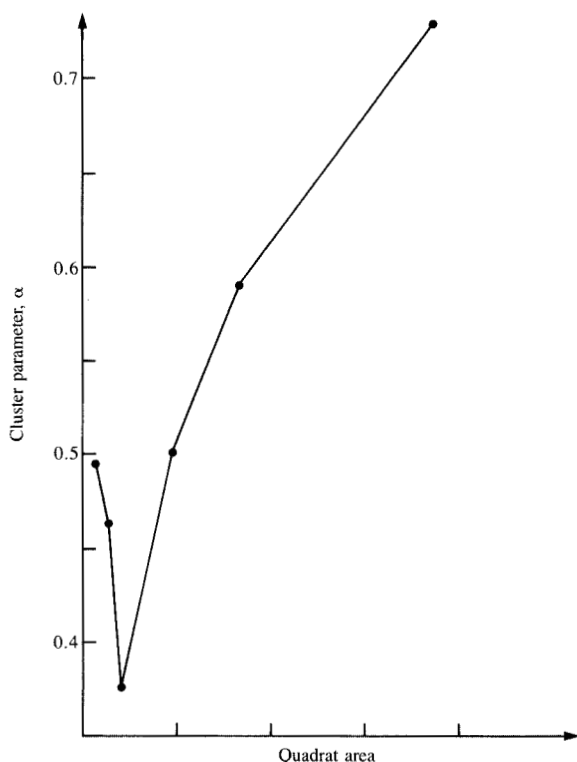


Figure 7

Clustering and quadrat area: The cluster parameter α as a function of the quadrat area.

parameter. With a nonlinear least-square technique, values of $Y_0 = 0.8339$, $D = 0.1479$, and $\alpha = 1.1772$ were obtained. When these values are used in formula (26), the yield-versus-area curve is the one indicated by the dashed line in Figure 8. The result fits comfortably between the 80% confidence limits of the data, suggesting that Equation (26) is an acceptable model. It is, furthermore, to be noted that these results were obtained for an area range from 1 to 36. This is believed to be the widest area range ever published for which this model is applicable.

Although Equation (26) does reasonably fit the data of Figure 8, it leaves us with a problem to resolve. The average particle density needed in Equation (26) was found to be 0.1479, while the results from Table 3 suggested a particle density of 0.4358 particles per unit area. Furthermore, in Figure 6 the values for α are seen to vary between 0.376 and 0.73, while for Equation (26) it is 1.1772.

The differences between the parameter values for Equation (26) and the data in Table 3 are caused by the use of the gross yield factor Y_0 . This quantity has been referred to in the industry as an "area yield factor" since it usually affects areas of wafers. A treatise on this can be found in the work by Ham, who called it the "area usage factor" [16]. Methods

for measuring this quantity have been described by Paz and Lawson [11], Warner [12], and Stapper [8]. In the work of Paz and Lawson the Y_0 factor contained the yield associated with clusters of defects, while Warner used it to take care of the extra yield losses in the regions of high defect density. The same is taking place in the data analyzed here. The Y_0 factor absorbs some of the yield of the clusters with high particle densities. As a consequence fewer particles are included in the defect density of Equation (26), thus resulting in a lower value for D . The value of α obtained for Equation (26) is higher than any of the values of α in Table 3. This indicates less clustering, which seems reasonable, since most of the high-density clusters are already included in the Y_0 factor.

The preceding results are a good illustration of the assumptions that are associated with yield models that incorporate a gross yield Y_0 or an area usage factor. The example also points out a hazard that could result from using such a factor. The average particle densities and the cluster parameter of the yield model are not necessarily the same as those of the actual particle or fault distributions. In general the particle densities or defect densities in such yield models tend to be lower than the ones observed experimentally. Similarly, the cluster parameters α are higher for yield-versus-area models than they are in the actual

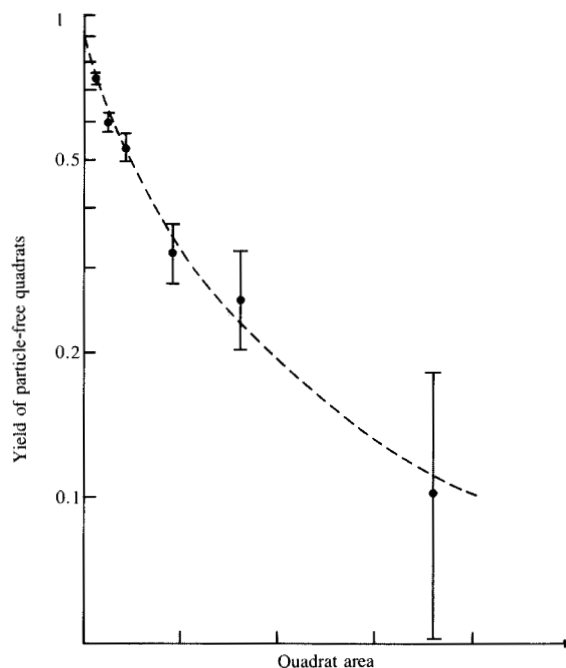


Figure 8

Yield and quadrat area: The yield-versus-area plot resulting from the maps of Figure 2. The bars indicate the 80% confidence limits.

distributions. These results are of prime importance in the modeling of yield with redundancy. In that application the correct defect densities and cluster parameters for the distributions are a necessity.

The difficulties resulting from the use of Equation (26) can be prevented by using the correct yield model. Since the frequency distributions of the number of particles per quadrat could be modeled with negative binomial distributions, the yield should be calculated with Equation (A5) of the Appendix. This formula can be written as

$$Y = [1 + AD/\alpha(A)]^{\alpha(A)}, \quad (27)$$

where the area dependence of the cluster parameter is indicated by $\alpha(A)$.

The values of α for use in Equation (27) are given in Table 3. Also shown in that table are the quadrat areas that can be used for A . With a particle density D equal to 0.4358 particles per unit area, the yields calculated with (27) are 0.732, 0.590, 0.522, 0.336, 0.225, and 0.103. The actual yields can be calculated from the distributions in Table 1. They are 0.738, 0.598, 0.530, 0.323, 0.259, and 0.104. These two results are sufficiently close for all practical purposes.

Obtaining the area dependence of the cluster parameter α in the preceding example was straightforward.

Unfortunately, it is virtually impossible to apply this technique to actual test data because the locations of all the defects and faults cannot be determined. An alternative approach is therefore required.

The two-step analysis of test data is often possible. In the first step the nature of the fault distributions has to be determined with the array diagnostics of [3], or by alternative fault diagnostic methods. The results should establish the model and parameters, like α , to be used for the fault distribution. The second step consists of doing a chip multiple or "window" analysis of wafer maps that show the locations of the functioning and defective chips. These results should establish the relationship between yield and chip multiples or area. The results from the two steps must then be combined to determine the area dependence of the parameters in the model. Since this requires the solution of transcendental equations, it is often necessary to use computer programs for calculating these parameters.

A simulation model

The data in the preceding sections came from only twelve wafers. It is difficult to ascertain whether the results and observations are unique for this sample, or whether they have wider implications. More data are needed to confirm whether the results are general. Unfortunately, the analysis by hand of particle maps is laborious and time-consuming. To avoid this, similar maps have been created with computer simulations. Some of these are shown in Figure 9. A large number of such maps have been analyzed with the quadrat method directly by computer.

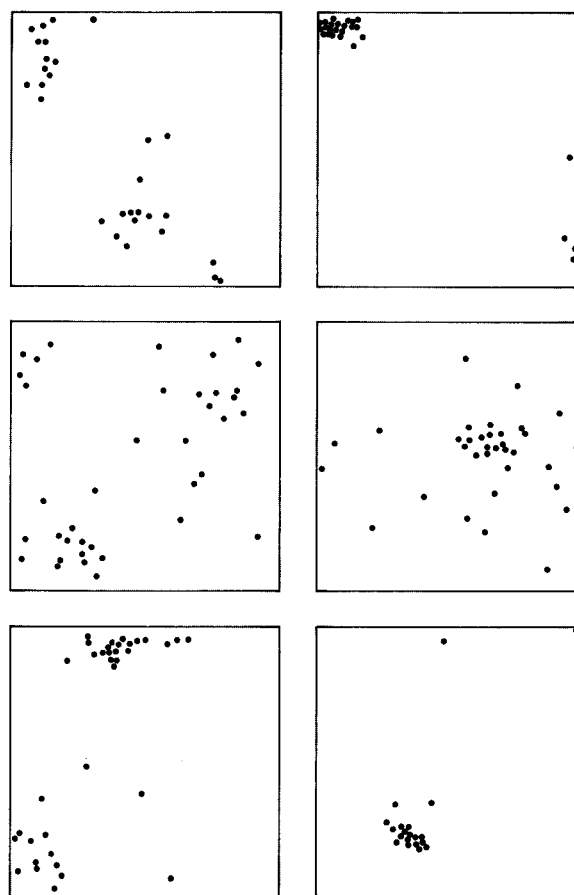


Figure 9

Simulated clusters: Cluster maps that were generated using computer simulations.

In the first step of the simulation process, clusters of hits that represent particles, defects, or faults are generated with a Gaussian probability distribution function. The probability for the occurrence of a hit is highest at the center of the cluster and decreases radially away from the center. The central probability and the standard deviation of each cluster were random numbers. The clusters were therefore all different and unique.

The hits or lack of hits within a cluster were determined with yet another set of random numbers. These numbers had values between zero and one and occurred at tightly spaced grid points. They were compared to the Gaussian probability calculated for each point. A hit was scored when such a number was smaller than or equal to this probability. The resulting clusters can be called symmetrical since they were generated with a radially symmetric probability distribution function. Many clusters of this type were stored in a cluster library for subsequent use. Clusters were taken at random from this library for further processing.

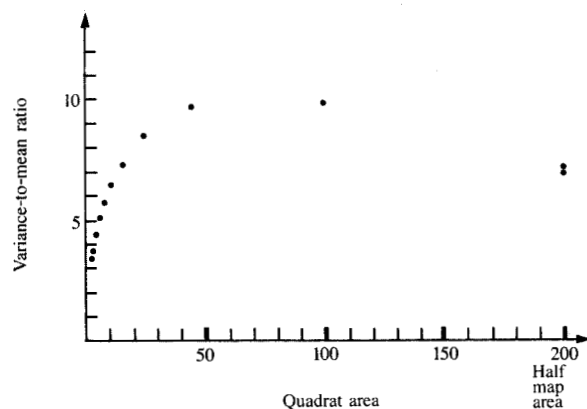


Figure 10

Variance-to-mean ratio: The deviation from Poisson statistics as a function of quadrat area for 200 simulated maps peaks at the size of 1/4 map.

Defect clusters on semiconductor wafers are rarely symmetrical. To simulate asymmetry, clusters from the library were projected onto wafer maps with randomly generated lateral and azimuthal angles. The centers of these clusters occurred at random spots on those maps. It was even possible for clusters to graze the edges of the maps, therefore causing most of the hits to fall outside the map area and leaving only a few inside. The number of hits required on a map was randomly selected from a predetermined probability distribution function. Clusters were thrown at the map until this number was reached or exceeded. If there were too many hits, the last cluster was randomly diluted to obtain the correct number.

Results from these simulated cluster maps have confirmed that the cluster parameter α varies as a function of quadrat area. In general, this dependence was smoother than the results in Figure 7. Until now, for all these simulation results, the minimum value of α occurred at the smallest quadrat size. More work still has to be done to determine the conditions that result in the minimum shown in Figure 7.

The simulations have also confirmed that the ratio of the variance to the mean is highest for the largest quadrat sizes, just as it did for the particles of Figure 2. It was possible to determine that this effect depends primarily on the distribution of the number of hits per map. This dependence has been studied with a special simulation experiment. In that experiment the map-to-map variation of the number of hits per map was forced to fit a Poisson distribution. A mean of 31.75 hits per map was chosen. This corresponded to an experimentally observed average number of particles per wafer.

Two hundred maps were generated and analyzed for this experiment. Each map consisted of an array of 420×420

points. All of these maps were analyzed with the quadrat method. The smallest quadrat size corresponded to a grid of 20×20 quadrats, while the largest quadrat size was equal to the complete map area. Areas of half maps were also analyzed. In that case horizontal and vertical partitioning of the maps resulted in two separate sets of data. The results of the experiment are plotted in Figure 10. The graph depicts the ratio of the variance to the mean as a function of area. This ratio reaches a maximum of 9.8 at an area that corresponds to four quadrats per map. The minimum of 1 is not shown on the graph. This occurs off-scale to the right for areas equal to the full map.

The maximum value of the ratio of the variance and the mean in Figure 10 is of great interest. This ratio is smaller than the ones observed in the map-to-map and wafer-to-wafer variations that were discussed earlier. For example, the distribution of the number of particles per map in Figure 2 had a variance-to-mean ratio of 28.3. Furthermore, particle counts on wafers in one factory resulted in an average of 31.75 particles per wafer and a variance of 4116.2. The ratio of the variance and the mean in that case is equal to 129.6. This is an order of magnitude higher than the maximum value resulting from the simulation experiment.

It therefore appears that, for the particle data, the map-to-map variations of the number of particles per map are the major source for the deviations from Poisson statistics. This suggests that a similar result can be expected from large variations in the numbers of defects or faults per wafer. Such variations could therefore be the cause of the increasing deviations from Poisson statistics observed in large and complex semiconductor chips.

Discussion and conclusions

Two effects have been described in this paper. One of these deals with the nature of the distributions of the number of particles, defects, or faults on integrated circuit chips. Such distributions appear to deviate increasingly further from Poisson's distribution when the integrated circuit complexity and chip area are increased. This effect has been studied with particle maps and with clusters that were simulated with a computer. Although the effect seems to be partially caused by clusters on wafers, it originates predominantly from the wide wafer-to-wafer variation in the number of particles, defects, or faults per wafer.

The second effect described in this paper has to do with the cluster parameter α associated with negative binomial distributions. This parameter has until now been treated as a constant. It was found that it actually varies with chip or quadrat areas when clusters are present.

Although these results dealt with negative binomial statistics, the area dependence of distributional parameters can be extended to other compound statistics. A number of such formulas are given in the Appendix. The cluster parameter α in any of these formulas can be treated as a

function of area as well. No matter which statistics are used, this area dependence of the cluster parameter affects the plot of yield as a function of area. Such plots could curve upwards or downwards or be bumpy, depending on the nature of clusters and the dependency of α on area.

With these possibilities in mind, it is timely to discuss one fallacy that has troubled yield modeling for some time. This has to do with data that show a linear relationship between the logarithm of the yield and the area of chips or quadrats. Such results are often believed to imply that pure, simple Poisson statistics are applicable as yield models. This, however, is not necessarily correct. The same results can be obtained with any of the yield formulas in the Appendix when α varies appropriately with area. A Poisson yield model is applicable only when the number of particles, defects, or faults per chip or quadrat are distributed according to Poisson's distribution. In semiconductor manufacturing this has not yet been the case, and according to the findings of this paper this is not ever expected to happen.

The dimensions of the photolithographic patterns also affect the fault distributions. As the integrated circuit industry matures, these dimensions are continually decreased. The result is an increased sensitivity to photolithographic defects. This has the same effect on the fault distributions as the increase in area had on particle distributions. The reduction in ground rules therefore causes an increasing deviation from Poisson statistics.

Another source of integrated circuit failures is the so-called pinholes that occur in the dielectrics that serve as insulation layers. These pinhole defects can be caused by flaws in the dielectric material or by particulate contamination in the manufacturing process. In either case the defect sensitivity increases if the dielectric thickness is decreased. Again, the trend in the semiconductor industry is towards thinner dielectrics and therefore towards increased defect sensitivities. An increased deviation from Poisson's statistics can therefore be expected to result from these defects also.

The faults caused by pinholes, photolithographic defects, and other types of defects have to be combined to obtain the fault distribution of a semiconductor product. This is done by the convolution of the distributions that are associated with each individual defect type. The actual statistics of defects and faults in integrated circuit chips are therefore quite complicated. The distributions in Figure 1 are an example of that. The long tails in those distributions are the result of increased defect sensitivities in a number of defect types. In the future the tails of such distributions are expected to become even longer.

Appendix

Negative binomial statistics as applied to clustering can be derived in a number of ways. At least four methods have

been described by Rogers [14]. The most commonly used derivation makes use of the compounding method of Equation (11). The compounder used in that case is a gamma distribution. In whichever way the negative binomial distribution is derived, the result can be written as

$$P(X = k) = \frac{\Gamma(\alpha + k)}{k! \Gamma(\alpha)} \frac{(AD/\alpha)^k}{(1 + AD/\alpha)^{\alpha+k}}, \quad (\text{A1})$$

where X is a random variable designating the number of faults, defects, or particles occurring in a chip, circuit, or quadrat with area A . The variable X can be equal to any positive integer value $k = 0, 1, 2, 3, \dots$, etc. The cluster parameter α in Equation (A1) must always be larger than zero.

The mean and variance of the negative binomial distribution are given by

$$E(X) = AD, \quad (\text{A2})$$

$$V(X) = AD(1 + AD/\alpha). \quad (\text{A3})$$

The distribution-generating function and the yield are given by

$$G(s) = [1 + (1 - s)AD/\alpha]^{-\alpha}, \quad (\text{A4})$$

$$Y = (1 + AD/\alpha)^{-\alpha}, \quad (\text{A5})$$

Although negative binomial statistics have shown very good agreement with integrated circuit data, they are not necessarily the only statistics applicable to integrated circuit manufacturing. Other compound Poisson statistics are available. One example of this is Neymann Type A statistics, which can also be derived in a number of ways. The most straightforward approach uses a Poisson distribution for the discrete compounder in Equation (7). The resulting formula has the form

$$P(X = k) = \frac{e^{-\alpha}(AD/\alpha)^k}{k!} \sum_{i=0}^{\infty} \frac{i^k}{i!} \alpha^i e^{-iAD/\alpha}. \quad (\text{A6})$$

The mean and variance of this distribution are

$$E(X) = AD, \quad (\text{A7})$$

$$V(X) = AD(1 + AD/\alpha). \quad (\text{A8})$$

The distribution-generating function and the yield associated with this distribution are

$$G(s) = \exp \{-\alpha[1 - e^{-(1-s)AD/\alpha}]\}, \quad (\text{A9})$$

$$Y = \exp [-\alpha(1 - e^{-AD/\alpha})]. \quad (\text{A10})$$

The method that has been used in this paper for generating compound distributions is not the only one available. Some of the distributions investigated by Armstrong and Saji are known as generalized Poisson distributions. They are derived by compounding a different Poisson equation than the one in (3) with a distribution that is the result of a complex convolution process. The Poisson-

binomial distribution is an example of this. It has the form

$$P(X = k) = e^{-\alpha} \sum_{i=0}^{\infty} \frac{\alpha^i}{i!} \frac{(ni)!}{k!(ni-k)!} \left(\frac{AD}{\alpha n}\right)^k \left(1 - \frac{AD}{\alpha n}\right)^{ni-k}, \quad (\text{A11})$$

where $n > 0$ and $0 \leq AD/\alpha n \leq 1$. The mean and variance of this distribution are given by

$$E(X) = AD, \quad (\text{A12})$$

$$V(X) = AD[1 + (n-1)AD/\alpha n]. \quad (\text{A13})$$

The formulas for the distribution-generating function and the yield are given by

$$G(s) = \exp[-\alpha\{1 - [1 - (1-s)AD/\alpha n]^n\}], \quad (\text{A14})$$

$$Y = \exp[-\alpha\{1 - (1 - AD/\alpha n)^n\}]. \quad (\text{A15})$$

The Poisson-negative binomial distribution is another example of a generalized Poisson distribution. It can be expressed by the formula

$$P(X = k) = e^{-\alpha} \sum_{i=0}^{\infty} \frac{\alpha^i}{i!} \frac{(ki+k-1)!}{k!(ki-1)!} \left(\frac{AD}{\alpha k}\right)^k (1 + AD/\alpha k)^{-ki-k}, \quad (\text{A16})$$

where $k > 0$ and $0 < AD/\alpha k < 1$. The mean and the variance of this distribution are given by

$$E(X) = AD, \quad (\text{A17})$$

$$V(X) = AD[1 + (k-1)AD/\alpha k]. \quad (\text{A18})$$

The distribution-generating function and the yield are

$$G(s) = \exp[-\alpha\{1 - [1 + (1-s)AD/\alpha k]^{-k}\}], \quad (\text{A19})$$

$$Y = \exp[-\alpha\{1 - (1 + AD/\alpha k)^{-k}\}]. \quad (\text{A20})$$

For all the preceding distributions the average number of particles, defects, or faults is given by AD and the cluster parameter by α . This cluster parameter has values between zero and infinity, or $0 \leq \alpha \leq \infty$. For $\alpha = \infty$ all these distributions become equal to the Poisson distribution, while $\alpha = 0$ signifies the case of maximum clustering. For this last condition all particles, defects, or faults are located in one small area, while all other surface areas are free from particles.

References

1. B. T. Murphy, "Cost-Size Optima of Monolithic Integrated Circuits," *Proc. IEEE* **52**, 1537-1545 (December 1964).
2. C. H. Stapper, "Modeling Redundancy in 64K to 16Mb DRAMs," *Digest of Technical Papers*, 1983 International Solid-State Circuits Conference, February 1983, pp. 86-87.
3. P. Gangatirkar, R. D. Presson, and L. G. Rosner, "Test/Characterization Procedures for High Density Silicon RAMs," *Digest of Technical Papers*, 1982 IEEE International Solid-State Circuits Conference, February 1982, pp. 62-63.
4. R. B. Seeds, "Yield, Economic, and Logistical Models for Complex Digital Arrays," *1967 IEEE International Convention Record*, April 1967, Pt. 6, pp. 60-61.

5. R. B. Seeds, "Yield and Cost Analysis of Bipolar LSI," presented at the 1967 International Electron Device Meeting, Keynote Session, October 1967 (Abstract p. 12 of the meeting record).
6. G. E. Moore, "What Level of LSI Is Best for You?," *Electronics* **43**, 126-130 (February 16, 1970).
7. R. M. Warner, "Applying a Composite Model to the IC Yield Problem," *IEEE J. Solid-State Circuits* **SC-9**, 86-95 (June 1974).
8. C. H. Stapper, "The Effects of Wafer to Wafer Defect Density Variations on Integrated Circuit Defect and Fault Distributions," *IBM J. Res. Develop.* **29**, 87-97 (January 1985).
9. C. H. Stapper, "On a Composite Model to the IC Yield Problem," *IEEE J. Solid-State Circuits* **SC-10**, 537-539 (December 1975).
10. A. P. Turley and D. S. Herman, "LSI Yield Projections Based Upon Test Pattern Results: An Application to Multilevel Metal Structures," *IEEE Trans. Parts, Hybrids, Packag.* **PHP-10**, 230-234 (December 1974).
11. O. Paz and T. R. Lawson, Jr., "Modification of Poisson Statistics: Modeling Defects Induced by Diffusion," *IEEE J. Solid-State Circuits* **SC-12**, 540-546 (October 1977).
12. R. M. Warner, "A Note on IC Yield Statistics," *Solid-State Electron.* **24**, 1045-1047 (December 1981).
13. F. M. Armstrong and K. Saji, IBM Development Laboratory, East Fishkill, NY, unpublished work.
14. A. Rogers, *Statistical Analysis of Spatial Dispersion*, Pion Ltd., London, 1974.
15. C. H. Stapper, "Yield Model for Fault Clusters Within Integrated Circuits," *IBM J. Res. Develop.* **28**, 636-640 (September 1984).
16. W. E. Ham, "Yield-Area Analysis: Part I—A Diagnostic Tool for Fundamental Integrated Circuit Process Problems," *RCA Rev.* **39**, 231-249 (June 1978).

Received September 5, 1985; accepted for publication December 11, 1985

Charles H. Stapper IBM General Technology Division, Burlington facility, Essex Junction, Vermont 05452. Dr. Stapper received his B.S. and M.S. in electrical engineering from the Massachusetts Institute of Technology in 1959 and 1960. After completion of these studies, he joined IBM at the Poughkeepsie, New York, development laboratory, where he worked on magnetic recording and the application of tunnel diodes, magnetic thin films, electron beams, and lasers for digital memories. From 1965 to 1967, he studied at the University of Minnesota on an IBM fellowship. Upon receiving his Ph.D. in 1967, he joined the development laboratory in Essex Junction. His work there included magnetic thin-film array development, magnetic bubble testing and device theory, and bipolar and field effect transistor device theory. He is now a senior engineer in the development laboratory, where he works on mathematical models for yield and reliability management. Dr. Stapper is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.